



Analysen zum Regressions-Index von santésuisse zur Wirtschaftlichkeitsprüfung bei Ärztinnen und Ärzte in der Schweiz

Michel Romanens, VEMS; Dr. Walter Warmuth, Gesundheitsforen Leipzig

Im Februar 2021

Copyright:

Nur unter Verwendung der Quellenangaben

WZWRegressionsindex022021

Weitere Informationen:

Educational slides:

WZW-Verfahren: <https://docfind.ch/WZWP122019.pdf>

Verhalten von PVK Mitgliedern: <https://docfind.ch/WZWPVK122019.pdf>

Aufsichtsbeschwerde an das BAG: <https://docfind.ch/Aufsichtsbeschwerde-FairfondBAG122019.pdf>

Einleitung

Im Zweifel für den Angeklagten – dieser Rechtsgrundsatz gilt in den Verfahren zur Beurteilung der Wirtschaftlichkeit frei praktizierender Ärztinnen und Ärzte nicht: Liegen ihre Durchschnittskosten 30% und mehr über dem Schnitt, sind sie grundsätzlich schuldig, es sei denn, sie können ihre Unschuld beweisen. Auch der Grundsatz *Keine Strafe ohne Delikt* gilt hier nicht: Ohne dass ihnen auch nur ein einziges Vergehen auf der Einzelfallebene nachgewiesen wurde, werden sie der Überarztung für schuldig gesprochen, allein aus statistischen Gründen. Dabei wird ihnen ein Angebot unterbreitet, wie sie sich mit einem Vergleich aus dem Verfahren kaufen können, in manchen Fällen noch bevor dieses überhaupt eröffnet wurde. Die meisten gehen darauf ein, um wieder in Ruhe arbeiten zu können. Es fällt schwer, in dieser Handhabung der Wirtschaftlichkeitsverfahren keine Ähnlichkeit mit einer Schutzgelderpressung zu sehen. Die Rechtsexperten sind sich heute auch weitgehend einig: Das Verfahren verstößt gegen die Grundsätze unserer Rechtsstaatlichkeit. Dass es zur Senkung der Gesundheitskosten beiträgt, zweifeln etliche Versicherungsmathematiker an, was der Grund sein dürfte, weshalb Deutschland den Durchschnittskostenvergleich viel selektiver und mit ausreichendem Vorverdacht durchführt. In der Schweiz sorgen alte Seilschaften dafür, dass sich der Durchschnittskostenvergleich hartnäckig hält, flankiert von einer untätigen FMH, mit teilweise ruinösen Folgen für Ärztinnen und Ärzte und in der Folge auch für deren Patientinnen und Patienten.

Die Systemfehler

Den Versicherern sind die hohen Kosten der frei praktizierenden Ärztinnen und Ärzte ein Dorn im Auge. Die Systematik, mit welcher sie unnötige Behandlungen zu identifizieren vermeinen, hat allerdings einen Fehler, welchen die Gerichtsbarkeit jahrelang nicht beachtet hat. Die Effekte der medizinischen Therapien, welche alleine wirtschaftlich geprüft werden müssen, bleiben dabei nämlich unberücksichtigt: Bei gleichen therapeutischen Effekten sind billigere Ärztinnen und Ärzte wirtschaftlicher. Doch diese gesetzlich vorgeschriebene Kernfrage zur Prüfung der Wirtschaftlichkeit interessiert niemanden. Kostenfolgen teurer Therapien besonders kranker Personen werden endogenen Faktoren zugeschrieben, was bedeutet: Der Arzt ist selber schuld, wenn er teure Patientinnen und Patienten behandelt. Diverse «Verbesserungen» der Verfahren basieren auf falschen Grundannahmen und perpetuieren die Systemfehler weit über das Jahr 2020. Der Grund dafür: Die Systemfehler gestatten den Krankenkassen Rückforderungen von rund 50 Mio. Franken pro Jahr. Da die Krankenkassen in Eigenregie die Methoden der Prüfungen diktieren, dazu noch die Datenbanken liefern und die daraus resultierenden Kostenbetrachtungen, liegt die Macht der Wirtschaftlichkeitsprüfungen vollkommen in den Händen des privaten, durch staatliche Kontrolle nicht kontrollierbaren Krankenkassenverbandes *santésuisse*, welcher subsidiär öffentlich-rechtlich relevante Evaluationsstandards diktiert und sich dabei nicht in die Karten schauen lässt. Im Ergebnis resultiert der grösste Justizskandal, den die Schweiz je zugelassen, bisher aber nicht erkannt hat. Aus diesem Grund veröffentlichen wir hier die notwendigen Hintergrundinformationen, damit der wahre Zweck der Wirtschaftlichkeitsprüfungs-Industrie einer Überprüfung durch unabhängige Personen zugeführt und hoffentlich als das erkannt wird, was er ist: ein Betrug, an den Gerichten, an den Ärztinnen und Ärzten – und letztlich an den Patientinnen und Patienten.

Teil 1: Die Industrie hinter den WZW Verfahren

Das Prüfverfahren von santésuisse generiert jährlich rund 2'500 auffällige Ärztinnen und Ärzte, welche dann aufgefordert werden, die überhöhten Kosten zu erklären. Um teils ruinöse Regresse zu vermeiden, willigen viele Ärztinnen und Ärzte in Vergleiche ein, welche so santésuisse ca 50 Mio. Franken pro Jahr generieren. Hinzukommen Anwaltskosten von im Schnitt rund 15'000 Fr. pro Fall, womit nochmals 37.5 Mio Fr. pro Jahr anfallen. Daneben fallen noch Zusatzkosten durch Beratungen an (Trustcenters, Schutzgemeinschaft für Ärztinnen und Ärzte (<https://www.s-g-a.org>), welche zum Teil von früheren Mitarbeitern von santésuisse geführt werden, wie z.B. Kristof Nagy (<https://ueberarztung.ch>) oder Dr. Jürg Reust¹. Es ist also keine Überzeichnung, hier von einer veritablen WZW-Industrie zu reden. Dieser Industrie geht es umso besser, je mehr auffällige Ärztinnen und Ärzte ihre Methode generiert. Wie präzise die Methode die tatsächlichen Überärzter identifiziert, tritt dabei in den Hintergrund, denn allein das allgemeine Klima der Angst vor Regressen führt insgesamt zu restriktiverem Behandlungsverhalten². Es existieren also etliche wirtschaftliche Interessen, möglichst viele Ärztinnen und Ärzte mit Regressverfahren zu bedrohen.

Teil 2: Analyse des B,S,S Gutachtens Dr. Boris Kaiser³

All diese Vorwürfe sind santésuisse bestens bekannt. Um sich insbesondere gegenüber der FMH, die mit santésuisse im gleichen Boot sitzt, wenn es um die Weiterentwicklung der Prüfmethode handelt, beliebter zu machen, generiert santésuisse regelmässig externe Gutachten. Das B,S,S Gutachten von Dr. Kaiser ist ein solches. Es lieferte im Jahr 2016 die mathematisch-konzeptionelle Grundlage für das Polynomics-Gutachten, welches 2017 mit individuellen und aggregierten Daten der Krankenversicherer getestet wurde³. Wir haben im Jahr 2016 das B,S,S Gutachten durch Dr. Walter Warmuth, einen ausgewiesenen Versicherungsmathematiker der Gesundheitsforen Leipzig, beurteilen lassen⁴. Er stellt als Erstes einen Systemfehler fest, wonach die Effizienz einer Praxis losgelöst vom Effekt dieser Praxis betrachtet wird. Man schaut nur auf die Kosten und vergleicht diese mit den Kosten in einer Vergleichsgruppe. In dieser Effizienzanalyse interessiert nicht, ob der Patient in der Folge verstorben ist oder überlebt hat und ob etwa Folgekosten direkter oder indirekter Art dank der Behandlung vermieden wurden. Da die Kosten immer für einen Effekt eingesetzt werden und nie um die Effizienz zu steigern, ist dieser Ansatz der Gesundheitsökonomie fundamental falsch: «Eine Behandlung ist effizienter als eine andere Behandlung, wenn sie bei einem vergleichbaren Effekt kostengünstiger ausfällt». Weiter stellt Dr. Warmuth fest, dass Durchschnittskosten durch Mengenausweitungen gesenkt werden können, wodurch die Durchschnittskosten einer Arztgruppe sinken, was den Mittelwert bei normalen Ärztinnen und Ärzte (ohne Mengenausweitung) und mit Pareto-verteilten Kosten bei hochbedürftigen Patienten künstlich erhöht. Dies ist der zweite Fundamentalfehler. Ferner formuliert Dr. Warmuth die Grundbedingungen für Regressionsanalysen:

- I. Die abhängige Variable (Kosten) wird als normalverteilt modelliert.¹
- II. Für die abhängige normalverteilte Variable wird als Modell eine lineare Funktion unabhängiger Variabler und ein zufälliger Fehler, der für alle anderen „Einflüsse“ steht, gewählt.

- III. Die zufälligen Fehler müssen sich im Mittel aufheben.
- IV. Die unabhängigen Variablen sind nicht zufällig, eine neue Beobachtung zeigt in Bezug auf diese Variablen immer das gleiche Ergebnis.
- V. Die unabhängigen Variablen sind linear unabhängig.
- VI. Schwankungen sind homogen, die Streuungen sind gleich.
- VII. Schwankungen sind nicht autokorreliert, Schwankungen verschiedener Beobachtungen sind unkorreliert.

Dr. Warmuth stellt fest, dass wegen der Paretoverteilung der Kosten zuerst auf ihre Gültigkeit geprüft werden müssen, was bisher nie erfolgt ist. Dies entspricht dem dritten Fundamentalfehler. Der Anspruch von Dr. Boris Kaiser, mit seiner Methode die Wirtschaftlichkeitsprüfung vom Effekt der Kosten loszulösen und trotzdem eine Evaluation über die «Wirtschaftlichkeit medizinischer Leistungen» von Ärztinnen und Ärzte in der Schweiz durchzuführen, entbehrt der wissenschaftlichen Redlichkeit und verletzt mehrere der 27 Kriterien der Schweizerischen Evaluationsgesellschaft⁵.

Teil 3: Die Analysemethode: eine Analyse des Polynomics Gutachten

1.1.Hintergrunddaten

Kernpunkt des Gutachtens ist die Verbesserung der Wirtschaftlichkeitsbeurteilung gemäss internationalen Standards. Dazu schreiben die AutorInnen: «In der internationalen Fachliteratur wird bei allen Verfahren zum sogenannten «Physician Profiling» die Morbidität des Patientenstamms berücksichtigt. Auch für die Schweiz wurde eine solche Berücksichtigung wiederholt gefordert (Schwenkglenks 2010¹; Wasem, Lux und Dahl 2010²).» Beide Forderungen basieren auf Gutachten, die der VEMS finanziert hatte. Nur dank dem Druck des VEMS im Jahr 2010 wurden parlamentarische Schritte unternommen, santésuisse zur Verbesserung der Verfahren zu zwingen.

Das Polynomics-Gutachten wurde unter Mitarbeit von Polynomics (Trottmann, Telser) erstellt. Auftraggeber sind santésuisse, curafutura und die FMH, Danksagungen erfolgten an Prof. Dr. Stefan Boes (Direktor des Zentrums für Gesundheit, Politik und Ökonomie an der Universität Luzern, Leiter des SNF-Programms Swiss Learning Health Systems) für die sorgfältige Lektüre und exzellenten Kommentare zu einer früheren Version dieses Berichts. Marc Bill (santésuisse), Mirjam D'Angelo (santésuisse), Dr. med. Andreas Häfeli (FMH), Thomas Kessler (FMH), Dr. Philip Moline (NewIndex) und Anke Trittin (curafutura), deren fundierte Rückmeldungen zu wesentlichen Verbesserungen der Studie geführt haben.

1.2. Zusammenfassung der Mängel (Mängelliste)

1. Es werden keine Diagnosedaten verwendet.
2. Es werden keine Vorjahresdaten verwendet.
3. Es werden keine Effekt-Variablen verwendet (Outcome-Daten).
4. Es wird nicht grundsätzlich erkannt, dass Wirtschaftlichkeit nur im Kontext mit Outcome erfasst werden kann, wonach niedrigere Kosten bei glei-

chem Outcome wirtschaftlicher sind (Fundamentalfehler Nr. 1 in der Gesundheitsökonomie).

5. Die Kostendaten sind nicht normal verteilt.
6. Nachträgliche Normierungen der Kostenverteilung sind experimentell.
7. Die Aggregation zu AGG führt zu aufgeblähten R² und Pseudoerklärungsgehalt (80% statt korrekt 20%).
8. Hohe notwendige Versorgungskosten werden automatisch zu einem spezifischen Praxiseffekt zugeordnet, Systemfehler Nr. 1 («ungerechtfertigt strenge Beurteilung»).
9. Aggregierte PCG auf Arzzebene in AGG verkennen die Gesamtmorbidity der Patientinnen und Patienten, dazu ist eine Individualberechnung auf Patientenebene notwendig, Systemfehler 2.
10. PCG weisen eine sehr hohe Varianz der Kosten auf, damit steigt die Fehlerbreite in kleinen Kollektiven (einer Arztpraxis) enorm an. PCG sind für das Physician Profiling nicht validiert. Es fehlt die Berechnung eines Vertrauensintervalls für die PCG selbst.
11. Das untere Vertrauensintervall wird nur für AGG berechnet. Das Vertrauensintervall auf Arzzebene ($\hat{\alpha}$) wird wegen fehlenden Daten ungenügend berechnet und für die Beta2-Variablen erst gar nicht gerechnet, die Wissenschaftlichkeit ist wegen den Mängeln bei der Datenbeschaffung nicht gegeben (SEVAL Kriterium B6⁵).
12. Schwellenwerte führen zu Verzerrungen, deren Ausmass und Bedeutung völlig unklar bleiben.
13. Wenn santésuisse einen Regressionsindex kommuniziert, ist nicht klar, mit welchem Modell gerechnet wurden (wurde eine Winsorisierung verwendet, wurde der Unsicherheitsindikator berücksichtigt, falls nicht, wie hoch war er?).
14. Die Zahl auffälliger Ärztinnen und Ärzte liegt weiterhin bei knapp 20%, die Methode sollte jedoch lediglich einen Anfangsverdacht bei rund 2-3% der Ärztinnen und Ärzte generieren, um nicht selber unwirtschaftlich zu sein, Systemfehler 3.
15. Der Effekt der zweiten Stufe ist minimal und damit nur eine potentielle

Fehlerquelle, somit hier ein Effizienzproblem und eine Verletzung des SEVAL-Kriteriums B7⁵.

16. Bei Berechnungen auf mit nicht-aggregierten, individuellen Patientendaten bleiben die Modelle die Erklärung von rund 70% der Kosten schuldig, Systemfehler 4.
17. Dass mit aggregierten Daten statt 14% nur 11% auffällige Ärztinnen und Ärzte resultieren ist kein Argument dafür, auf die Individualdaten zu verzichten.
18. Die Übereinstimmung zwischen Berechnungen mit individuellen und aggregierten Daten ist bei einem Drittel der auffälligen Ärztinnen und Ärzte nicht vorgegeben, die Methode ist somit in sich widersprüchlich und damit nicht ausreichend validiert (SEVAL B8 verletzt⁵).
19. Die Autoren legen sich nicht fest, mit welcher Methode schliesslich gerechnet werden sollte. Damit kann santésuisse weiterhin willkürliche Inzidenzen von auffälligen Ärztinnen und Ärzte pro Jahr produzieren (18%). Allein schon die potentielle Anklage durch eine nicht Outcome-validierte Methode ist ehrverletzend und verletzen ethische Grundprinzipien (SEVAL Kriterium A10) und sind respektlos (SEVAL Kriterium A11).
20. Es gibt keine externe Zertifizierung betreffend die Frage, ob sasis/santésuisse richtig rechnen. Da santésuisse Partei in diesen Gutachten darstellt, haben wir hier einen weiteren Systemfehler, welcher auch schon vom kantonalen Ethikrat für Statistik moniert wurde⁸, zudem hat santésuisse nie die Charta dieses Ethikrats unterzeichnet.
21. Medikamentenkosten müssen aus den Wirtschaftlichkeitsverfahren eliminiert werden ?.

1.3. Weitere Ausführungen zur Mängelliste

Morbiditätsinformationen in Form von diagnostischen Informationen werden in der Literatur am häufigsten für das physician profiling eingesetzt (Kapitel 3.1.1 im Polynomics-Gutachten). Diese Informationen sind jedoch wegen fehlenden Daten seitens santésuisse und sasis nicht verfügbar. Die Vorjahres-Kosten sind ebenfalls ein guter Morbiditätsindikator, welcher aber wegen fehlenden Daten

ebenfalls nicht verwendet wird.

Die Wahl des Regressionsmodells (OLS, GLM, ecc) ist generell weniger wichtig als die Risiko-Adjusters (Eijkenaar und van Vliet, 2014). Insbesondere bei Modellen, welche wenig der Kosten erklären (niedriges R^2), muss von ungenügender Zahl erklärender Variablen ausgegangen werden.

Daraus folgt, dass das Profiling-Modell selber einen Fehlerindikator benötigt. Der Fehler ist umso grösser, je schlechter das Modell die Kosten erklärt. Das R^2 beinhaltet insbesondere bei der Individualdaten-Analyse meist weniger als 40%, entsprechend besteht ein Fehlerindikator von 60%. Die Test-Zuverlässigkeit wird aber hier mit der Varianz zwischen den Praxen geteilt durch die Varianz zwischen den Praxen und innerhalb einer Praxis. Dieser Zuverlässigkeitstest ist allerdings rein theoretisch und basiert nicht auf einer empirischen Prüfung. Dies bedeutet: die Fehlerindikatoren sind selber nicht validiert.

Das Kernstück der Analyse betrifft Alters- und Geschlechtsgruppen (AGG) erzeugte logarithmisch transformierte Effekte (β_1) plus den erklärenden Variablen (β_2) plus der praxisspezifischen Konstante (α) plus der stochastische Störterm (ϵ) pro Facharztgruppe in einem fixed effects Modell. Der zentrale Nachteil dieses Modells ist, dass indizierte Leistungen mit hohen Kosten verbunden sein können, welche dann fälschlicherweise einem spezifischen Praxiseffekt zugeordnet werden. Die Autoren belassen aber diesen Systemfehler, vertrauend darauf, dass in der weiteren Analyse der Praxis (analytische Methode) solche spezifischen Leistungsspektren ohnehin erkannt würden. Das Problem ist hier, ob sie auch als Entlastung anerkannt werden. Das Fixed effect Modell wird dann anhand einer zweiten Regression erweitert, wo fixe Praxiseffekte als abhängige Variable ($\hat{\alpha}$) berücksichtigt werden wie etwa kantonaler Standort der Praxis, die Facharztgruppe (FAG), ergänzt durch weitere Faktoren (Z) und ergänzt mit Schätzkonstante (ρ) und Störterm (u). Für dieses zweistufige Modell existiert jedoch kein Standard, es ist experimentell, und es resultieren Verzerrungseffekte, deren Ausmass im Einzelnen nicht genügend erforscht sind.

Der Unsicherheitsindikator für praxisspezifische Effekte ($\hat{\alpha}$), welcher auf den Residuen und Anzahl Beobachtungen pro Praxis basiert, wird auf der Ebene AGG gerechnet, je höher die Variabilität der Homogenität ist, desto höher ist der Unsicherheitsindikator aber nicht praxisspezifisch. Der Unsicherheitsindikator für die erklärenden Variablen (beta) wird jedoch nicht auch noch berechnet. Damit werden praxisunspezifische Kosten wie etwa die durchschnittliche Patient morbidity ohne Unsicherheitsindikator gerechnet. Auch für die enorme Varianz der Kosten pro PCG erfolgt keine Berechnung der Unsicherheit, ein zentraler Fehler bei der Einführung der PCG in das Physician Profiling. Die Berechnung des Unsicherheitsindikators erfolgt dann ausschliesslich für die Untergrenze, um die Zahl falsch positiver Praxen zu reduzieren und entspricht dem unteren 95% Konfidenzintervall. Dabei wird vom Punktschätzer der Unsicherheitsindikator abgezogen.

Um die Regressionsmodelle rechnen zu können, wird eine Normalverteilung vorausgesetzt. Berechnungen ohne logarithmische Transformation werden in der Regel bevorzugt. Diese Fragen sind praktisch ungelöst, das Vorgehen ist experimentell.

Erklärende Variablen mit ausreichender Exogenität (niedriges Risiko für Manipulationen) sind PCG, Diagnosen, Franchisen; weniger gut z.B. Hospitalisationen im Vorjahr. Praxen, die dank guter Medizin diagnose- und medikamentenvermeidende Wirkung haben, erscheinen jedoch dann als kostenauffällig («ungerechtfertigt strenge Beurteilung»), insbesondere also Praxen mit niedrigen PCG / Medikamentenkosten, wie etwa in der Alternativmedizin. Dies muss bei der inhaltlichen Interpretation und Modellbeurteilung beachtet werden.

Für die Analyse werden aggregierte PCG pro AGG verwendet, Individualdaten sind nicht verfügbar. Die PCG wurden nicht speziell für Wirtschaftlichkeitsverfahren entwickelt. Die Operationalisierung der PCG ist komplex (z.B. DDD) und Schwelleneffekte können problematisch werden. Bei sehr hohen Kosten erfolgt

eine Winsorisierung auf der 95 Perzentile. Damit wird die schiefe Verteilung von 6.45 auf 0.80 reduziert.

Der Erklärungsgehalt der AGG mit aggregierten Daten ist sehr hoch (70-90%), eine Folge der Aggregation und würde auf der Individualebene lediglich 10% betragen (das Problem ist hier, dass die Aggregation der Daten die Varianz so massiv reduziert, dass das R² künstlich aufgebläht wird).

Die AGG werden bei der Beurteilung des Erklärungsgehalts von Praxiseffekten (PE) und PCG (volles Modell) beibehalten, wodurch sich für alle Modelle ein massiv aufgeblähtes R² ergeben, welches das meiste an Erklärungsgehalt liefert (Tabelle 12). Entsprechend minim ist der zusätzliche Erklärungsgehalt durch PCG (um 1%!).

Gemäss Tabelle 17 sind die Resultate des errechneten Indexes ohne Winsorisierung unplausibel, weil Indexe bis zu 32'569% resultieren können. Die Anzahl Praxen mit einem Index über 130% ist denn auch mit Winsorisierung am niedrigsten und schwankt zwischen 9% (Kardiologie) und 27% (Psychiatrie und Psychotherapie).

Je höher der Index ist, desto höher ist der Unsicherheitsindikator. Gemäss Abbildung 7 ist es im Einzelfall möglich, dass die Untergrenze des Indexes 130% beträgt, tatsächlich aber ein Index von 250% ausgewiesen wird. Unter Verwendung des Unsicherheitsindikators beträgt der Anteil auffälliger Praxen statt 18% nur 11% und wird insbesondere in der Psychiatrie mehr als halbiert (Tabelle 22). Die zweite Stufe hat einen nur unwesentlichen Einfluss auf die Anzahl auffälliger Praxen, sie sinkt von 2'775 auf 2'661 Praxen (Tabelle 24). Die Zahl auffälliger Praxen beträgt insgesamt rund 18% von 15'766. Unter Verwendung von künstlich erzeugten ineffizienten Praxen (Tabelle 27) werden leider die Prävalenzzahlen verschwiegen, gemäss eigenen Berechnungen dürfte die Zahl auffälliger Praxen bei 22% liegen.

Eigene Berechnungen zu Tabelle 27:

	TP	TN	FP	FN	SENS	SPEC	PPV	NPV	ACC	ALL	pLR	nLR	PRE%
Poly-nomics	1504	13700	5	2255	40.0	99.96	99.7	85.9	87	17464	1096.7	0.60	22

Abbreviations: TP=true positive, TN=true negative, FP=false positive, FN=false negative, SENS=Sensitivity, SPEC=specificity, PPV=positive predictive value, NPV=negative predictive value, ACC=accuracy, pLR=positive likelihood ratio, nLR=negative likelihood ratio, PRE=prevalence

Auffällig ist der exzellente positiv prädiktive Wert für die M3 untransformierte Methode. Welche Methode nun aber präferabel ist, kann aufgrund der Tabelle 27 nicht abschliessend beurteilt werden. Gefordert sind aber möglichst spezifische Tests auf Kosten der Sensitivität. Im Übrigen ist die Prävalenz der auffälligen Praxen in Tabelle 27 fiktiv.

Für die Berechnungen mit Individualdaten werden 1'420 Praxen mit 369'894 Patientinnen und Patienten beigezogen. Dabei zeigt sich infolge der fehlenden AGG-Aggregationen ein echtes R² von 8% bis 24% für Alter und Geschlecht (statt der aufgeblähten 80%) und eine signifikante Verbesserung des R² auf rund 35% unter Verwendung der Morbiditätsvariablen in Tabelle 31. Der Erklärungsgehalt des Modells bleibt aber bei rund 70% der Kosten im Ungewissen. Ob es sich dabei um exogene oder endogene Variablen handelt, wird nicht fassbar. In der Kardiologie sind zudem die Kosten nur zu 16% erklärt (Psychiatrie 21%), und die Verwendung der PCG auf Arzzebene ändert kaum etwas (M3). Zentral von Bedeutung ist die Tabelle 33. Hier werden die Individual-Daten und die aggregierten Daten betreffend die Zahl ineffizienter Praxen direkt verglichen. Während mit den Individualdaten 14.1% auffällige Praxen resultieren, sind es mit den aggregierten Daten nur 10.7%.

Die Übereinstimmung ist für beide Methoden mit 9.4% angegeben, woraus sich ein Kappawert von 0.73 schätzen lässt. Die Übereinstimmung bei den auffälligen Praxen liegt bei 66%. Bei einem Drittel der Praxen liefern also die beiden Methoden divergente Resultate.

Unter Einbezug einer Sicherheitszone (Praxen mit Untergrenze) ergibt sich jedoch nochmals ein völlig anderes Bild. Der Kappawert bleibt zwar mit einem geschätzten Wert von 0.74 hoch, die Übereinstimmung betrifft nun jedoch nur noch 40% und die Prävalenz der auffälligen Praxen sinkt übereinstimmend in beiden Modellen auf 2.3%.

Eine klare Begründung für diese Effekte liefert das Gutachten nicht.

Die weiteren Simulationsberechnungen (Tabelle 34) sind nicht nachvollziehbar, da die Vierfelder-Tafel keine Zahlen für richtig positive und richtig negative liefert, womit auch die Prävalenzzahlen nicht eruiert sind und unklar bleibt, was genau berechnet wird. Unter Berücksichtigung von Unsicherheitsindikatoren beträgt die kongruente (für zwei Modelle berechnete) Prävalenz von Überarztung für rund 17'000 Ärztinnen und Ärzte ca 2.3%. Polynomics akzeptiert im Bericht eine Prävalenz von 18%. Damit beträgt die Zahl der potentiell falsch positiven Ärztinnen und Ärzte weiterhin 9:1.

Teil 4: Fragen zur Rechtskultur

Die Prüfung der Wirtschaftlichkeit mit einer statistischen Methode ist das einzige in der Schweiz zugelassene Beweismittel¹⁰. Dies bedeutet, dass die Erarbeitung von prognostischen Vorannahmen, welche allenfalls einen Beweis mitbegründen könnten, der Statistik von santésuisse überlassen werden, obwohl santésuisse grosses Interesse daran hat, auffällige Ärztinnen und Ärzte zu produzieren, wie der Ethikrat für Statistik der Schweiz richtig bemängelt hat⁸. Während bei der Beurteilung einer Straftat des Strafgerichts, wie etwa bei einem Mordfall, gewisse Vorannahmen richterliche Entscheide beeinflussen, berücksichtigen unsere Versicherungsgerichte bei der Frage der Überarztung allerdings keine Vorannahmen. Solche wären etwa eine Verschuldung oder die Anschaffung teurer, zu amortisierender Geräte etc. Dies ist zu bemängeln, liefern solche Vorannahmen doch mitunter wichtige Hinweise auf eine Motivation zur Überarztung, etwa wenn eine Praxis Geräte anschafft, die mit ihrem Patientengut und in ihrem Einzugsgebiet ohne Überarztung schwerlich zu amortisieren sein dürften. Stattdessen wird eine grundsätzliche Motivation angenommen: die Motivation der Ärztinnen und Ärzte, ihren Umsatz zu steigern. Dies mag ökonomisch betrachtet einleuchten, hat aber den Charakter einer generellen Vorverurteilung und widerspricht insofern der Sicht der Unvoreingenommenheit, zu der das Recht verpflichtet ist¹¹.

Im Ergebnis entsteht bei erhöhten Kosten also die Situation, dass eine Beweisumkehr stattfindet, indem Ärztinnen und Ärzte gegenüber santésuisse die erhöhten Kosten begründen müssen, ohne dass dies regelrecht über Outcome-Daten erfolgen würde. Die Folge davon ist, dass Ärztinnen und Ärzte die erhöhten Kosten nicht begründen können und dann rasch in Vergleichszahlungen einwilligen. Dabei kann es vorkommen, dass santésuisse für die folgenden Jahre gleich auch noch Regresse einfordert mit der Begründung, man wolle

sich ja nicht jedes Jahr wieder zu Vergleichsverhandlungen treffen. Zudem wurden und werden die Medikamentenkosten weiterhin falsch bewertet. Prof. Kieser hat hierzu für den VEMS ein Kurzgutachten⁹ erstellt und stellt fest, dass bei der Abklärung der Wirtschaftlichkeit ärztlicher Leistungen die Medikamentenkosten anders als andere ärztliche Leistungen und Kosten zu berücksichtigen sind. Medikamente sind durch die behördlichen Verfahren also per Definition WZW-konform und müssen entsprechend in den Berechnungen der Wirtschaftlichkeit (RSS-Index, Anova-Index, Regressions-Index) als Morbiditäts-Variable verwendet werden.

Im Ergebnis führt dies dazu, dass der Gesamtkosten-Index durch den Medikamentenkosten-Index geteilt werden muss. Diese einfache aber sehr wirksame Morbiditätskorrektur wurde von uns in einem Artikel in der Schweizerischen Ärztezeitung aufgearbeitet und erklärt². Der VEMS hat diesen Systemfehler bei der Beurteilung der Wirtschaftlichkeit entdeckt und durch oben erwähntes Kurzgutachten von Prof. Kieser rechtswissenschaftlich bestätigen lassen. Damit sind sämtliche Urteile auf Ebene PVK, Kanton und Bund zu den Wirtschaftlichkeitsverfahren, welche diese Korrektur nicht berücksichtigt haben, obsolet. Nun ist es an der Medizin, auf diese klare Rechtslage entschlossen zu reagieren und zu Unrecht erhobene Regresse zurückzufordern. Insbesondere sind diese Fakten in die laufenden Verfahren einzubringen, die teilweise noch immer auf falscher Beurteilungsgrundlage von santésuisse erbittert weiterverfolgt werden, in manchen Fällen sogar dann, wenn die Anschuldigung zweifelsfrei falsch ist und ihre Falschheit von den Inspektoren von santésuisse sogar zugegeben wird.

Auf der Ebene der paritätischen Vertrauenskommission, welche angeschuldigte Ärztinnen und Ärzte in gewissen Kantonen anhört, existieren aktuell keine weiteren Entwicklungen, insbesondere sind die anwesenden Experten, welche nicht von den Krankenkassen gestellt werden, kaum ausreichend statistisch, ethisch und mathematisch geschult, um den Anfangsverdacht einer kritischen Prüfung zu unterziehen. Häufig bieten diese «Experten» Hand dazu, die Höhe der Vergleichszahlungen zu reduzieren, um so ihre Anwesenheit moralisch zu

rechtfertigen. Grundsätzlichere Fragen zu den Daten von santésuisse werden dabei nicht diskutiert, sondern schicksalhaft hingenommen.

Teil 5: Der Regressions-Index als Medizinprodukt

Medizinprodukte sind rechtlich geregelt (<https://www.fedlex.admin.ch/eli/cc/2001/520/de>) und beinhalten Instrumente zur Erkennung und Verhütung von Krankheiten, z.B. mit Software im Sinne eines klassischen Medizinproduktes.

Ob ein Medizinprodukt ein Medizinprodukt ist, wird durch die Bestimmung definiert. Jedes Produkt, welches Einfluss auf gesundheitliche Belange und Entschiede deklariert, ist ein Medizinprodukt. Es genügt demnach der Nachweis, dass eine Einflussnahme durch Medizinprodukte auf medizinische Belange erfolgt, um den Nachweis zu erbringen, dass ein Produkt ein Medizinprodukt ist. Dabei ist es unerheblich, ob dieses Produkt ein Sachprodukt ist oder nicht. Es handelt sich in diesem Fall um ein Dienstleistungsprodukt.

Sind Medizinprodukte problematisch, indem sie beispielsweise die Gesundheit gefährden können, sind die Produzenten darauf hinzuweisen. Kommen diese den gewünschten Anpassungen nicht nach, kann als nächstes die Behörde, welche für die Überwachung der Medizinprodukte zuständig ist, kontaktiert werden. In der Schweiz erfolgt dies mit einer Anzeige an SwissMedic.

Der Regressions-Index von santésuisse ist ein Medizinprodukt der Klasse I und unterliegt den gesetzlichen Rahmenbedingungen, Art. 86 und Art. 87 HMG umschreiben die strafbaren Tatbestände. Aufgrund der vorliegenden Analyse ist davon auszugehen, dass Patientinnen und Patienten durch den Regressions-Index geschädigt werden, insbesondere durch unterlassene medizinisch notwendige Behandlungen.

Teil 6: Schlussfolgerungen und Ausblick

Unsere Analysen belegen, dass die Wirtschaftlichkeitsverfahren systematisch fehlerhaft angeblich unwirtschaftliches Verhalten detektieren und damit eine massive Belastung des ordentlichen Funktionierens der ambulanten Medizin bewirken. Der Anspruch der Wirtschaftlichkeitsverfahren, Kosten zu senken, resultiert im Gegenteil: Overhead-Kosten ungeahnten Ausmasses, Burnout bei korrekt arbeitenden Ärztinnen und Ärzte, Bedienung von Pseudonarrativen («Abzocker-Ärzte») und Gewinnoptimierung bei den Krankenversicherern.

Santésuisse hat es geschafft, faktisch eine Blanko-Methode zu bewirtschaften, ohne dass öffentlich-rechtliche Kontrollen stattfinden. Dies ist einem Systemfehler der Politik seit 100 Jahren geschuldet, dass nämlich die Krankenkassen die Beweis-Hoheit über Unwirtschaftlichkeit erhalten haben. Dass santésuisse diese Beweis-Hoheit in ihrem Interesse gestaltet hat und durch die Gerichte sanktionieren liess, wirft ein Schlaglicht auf die statistische Überforderung und Unprofessionalität der Gerichtsbarkeit, wie sie von Altwicker exakt herausgearbeitet wurde¹⁰. Diese Einschätzungen sind nicht die unseren, sondern Erzeugnisse rechtswissenschaftlicher Institute. Das sollte hellhörig machen, sich mal wirklich unvoreingenommen mit der Thematik zu beschäftigen.

Literatur

1. Reust JB. Wirtschaftlichkeitskontrollen Neuerungen Fallen-worauf ist zu achten? Hotel Wilder Mann, Luzern [Internet]. 2018 [cited 2019 Nov 10]. Available from: www.jbreust.ch
2. Romanens M, Warmuth W, Schober EA, Koop P, Kurth F. Der MBI zeigt , ob Sie richtig - beurteilt werden. *Schweizerische Aerztezeitung*. 2019;100:1401–1403.
3. Kaiser B. Methodische Weiterentwicklung der Wirtschaftlichkeitsprüfung. 2016; Available from: <https://docfind.ch/WZWEmpfehlungBSS032016.pdf>
4. Warmuth W. Stellungnahme zu Methodische Weiterentwicklung der Wirtschaftlichkeitsprüfung. 2016; Available from: <https://docfind.ch/WZWGutachtenWarmuth032016.pdf>
5. Rüefli C, Zweidler R, Koci M. Evaluationsstandards der Schweizerischen Evaluationsgesellschaft (SEVAL) [Internet]. 2016 [cited 2019 Nov 10]. Available from: https://www.seval.ch/app/uploads/2018/01/SEVAL-Standards-2016_d.pdf
6. Schwenkglens M. RSS versus PSS. Available from: www.physicianprofiling.ch/rsspssstatistikbericht052010.pdf
7. Lux G, Dahl H, Wasem J. Beurteilung des Screening-Verfahrens der Krankenversicherer in der Schweiz zur Identifikation von Überarztung Bericht zu einem vom Verein Ethik und Medizin Schweiz geförderten Projekt. *Univ Duisbg Ess* [Internet]. 2010; Available from: <https://www.wiwi.uni-due.de/fileadmin/fileupload/WIWI/pdf/182.pdf>
8. Tschopp P. Stellungnahme des Ethikrates zur Leistungserbringerstatistik von santésuisse. [Internet]. 2005; Available from: http://www.ethikrat-stat.ch/de/assets/File/faelle/a1_Scan.pdf
9. Kieser U. Berücksichtigung von Medikamentenkosten bei der Wirtschaftlichkeitsprüfung [Internet]. 2018. Available from: <https://docfind.ch/Kieser082018.pdf>
10. Altwicker T, Biaggini G, Marti A, Widmer L, Karlen P, Auer C. Statistikbasierte Argumentation im Verwaltungsrecht. *Schweizerisches Zentralblatt für Staats- und Verwaltungsr.* 2018;1–17.

11. Romanens M, Warmuth W, Schober T, Koop P, Kurth F. WZW Aufbereitung [Internet]. 2019. Available from: <https://www.docfind.ch/VEMSAufarbeitungWZW.pdf>